

What's In a Name? Issues in Named Entity Recognition

Colleen Kron, William L. Little, James C. Wolfe, Petra Ajaka, Benjamin O. Allen,
Christopher Brown, Marie-Catherine H. de Marneffe, Micha Elsner, Alexander H
Erdmann, Matias D. Grioni, Brian D. Joseph, Andrew B. Kessler, Hannah F. Young

THE HERODOTOS PROJECT: *a digital humanities project aiming to develop a virtual museum of information about groups of people (“ethnic” groups) recognized by the ancients as significant, with the compilation of a comprehensive catalogue of such peoples as the first step (<https://u.osu.edu/herodotos/>)*

0. GOALS FOR PRESENTATION

- 0.1 Introduce OSU's *Herodotos Project* as it pertains to onomastics
- 0.2 Discuss issues the project has faced that are of a general, somewhat theoretical, nature for onomastic research

I. NAMED ENTITY RECOGNITION (NER)

- 1.1 NER = a machine learning technique for automatically identifying “named” entities” in a text or discourse, where NAMED means a referential entity using a name and ENTITY here simply means something that occurs in a text
- 1.2 NER requires training on large amounts of text in which humans have manually, accurately, and consistently identified the named entities of interest
- 1.3 Computational algorithm takes over after training on the annotated text – picks out the named entities of interest in non-annotated text
- 1.4 How we annotate: BIO style annotation (Beginning, Inside, Outside)

Personal Names	/PRS
Group Names	/GRP
Geographic Locations	/GEO
Null	/0

II. ONOMASTIC ISSUES: What *IS* a name?

- 2.1 Consistency or Context
Ambiguity
Morphology
- 2.2 Ultimately, every noun can be construed as a name
ex. forum/0
forum/GEOB Romanum/GEOI

mons/0
mons/GEOB Vesuvius/GEOI

III. THREE CASES OF NAMING AMBIGUITY IN GREEK (HERODOTOS, *HISTORIAE*)

3.1 THE DELPHIANS (Distinguished by Preposition)

- 3.1.1 The Greek word **Δελφοί, ὧν** has several related meanings: “the oracle located at Delphi”, “the place Delphi,” and “the people who live in the place Delphi.”

Δελφοί, ὧν, οἱ, Delphi, a famous oracle of Apollo in Phocis at the foot of Parnassus, once called Pytho, as in Il. 9. 405, and always in Hdt., as I. 54; called Delphi first in h. Hom. 27. 14, Soph. O. T. 734. **II. the Delphians**, Hdt. I. 54, etc.: also in sing., **Δελφός** as king of Delphi, Aesch. Eum. 16; **Δελφός ἀνὴρ** Eur. Andr. 1151, etc.: fem. **Δελφίς** Soph. O. T. 463, etc.; Adj. **Δελφικός, ἡ, ὄν, Delphic, Delphian**, Id. O. C. 413, Plat., etc.

Liddell & Scott, *Greek English Dictionary*

- 3.1.2 Agency = GRP

ὑπὸ Δελφῶν καλέεται Γυγάδας (Hdt.1.14)
*(it) is called “Gygian” **by the Delphians**.*

Location = GEO

ἐν δὲ Δελφοῖσι ὡς ἐσῆλθον τάχιστα ἐς τὸ μέγαρον (Hdt.1.47)
*But **at Delphi**, as soon as they went into the hall...*

Ambiguous = ??

ὁ δὲ ὡς τὸ **ἐκ Δελφῶν** ἤκουσε (Hdt.1.48)
*But when he heard (this) **from the Delphians**.*

3.2 THE LYDIANS (Distinguished by Number)

- 3.2.1 Singular (**ὁ Λυδός**) is commonly used to refer to Croesus “The Lydian”

Plural (**οἱ Λυδοὶ**) is not multiple Croesuses, but is used to refer to “The Lydians” as a group.

ὡς ἐσῆλθον τάχιστα ἐς τὸ μέγαρον **οἱ Λυδοὶ** χρησόμενοι τῷ θεῷ
καὶ ἐπειρώτων τὸ ἐντεταλμένον...

*As soon as **the Lydians** went into the hall they consulted the god
and asked what they were ordered to ask...*

3.3 THE PYTHIA (Distinguished by Historical Context)

- 3.3.1 ἡ Πυθίη = “The Pythia” = the oracle

- 3.3.2 Not like Croesus “The Lydian”. It is, as far as we can see, only ever used to refer to the oracle at Delphi

ἡ Πυθίη ἐν ἑξαμέτρῳ τόνῳ λέγει τάδε. (Hdt. 1.47)
The Pythia says the following in hexameter verse.

ταῦτα οἱ Λυδοὶ θεσπισάσης τῆς Πυθίης συγγραψάμενοι οἴχοντο
ἀπιόντες ἐς τὰς Σάρδις. (Hdt. 1.48)
*The Lydians, after the Pythian gave her prophecy, wrote down
these things, departed, and went back into Sardis.*

IV. TWO CASES OF NAMING AMBIGUITY IN LATIN (Pliny the Elder, *Historia Naturalis*)

4.1 JUNOS (Authorial Intention and Tone)

4.1.1 Does the “tone” of the author change how we understand his use of the plural for “Junos”

cum singuli quoque ex semet ipsis totidem deos faciant Iunones
Geniosque adoptando sibi, gentes vero quaedam animalia et aliqua etiam
obscena pro dis habeant ac multa dictu magis pudenda (Pliny the Elder *NH*
2.16.6)

*Since everyone makes as many gods as he pleases, adopting for himself
Junos and Geniuses. There are some nations that have animals and even
some obscene things as gods, and even many more shameful to say.*

4.2 SCIPIO AFRICANUS (Overall Project Goals)

4.2.1 We are mining names for names of groups – we are not creating a prosopography!

auspicatius e necta parente gignuntur, sicut Scipio Africanus prior natus
primusque Caesarum a caeso matris utero dictus, qua de causa et Caesones
appellati. (Pliny the Elder *NH* 7.47.1)

*It is more auspicious when the mother dies in giving birth, such was the
case for the birth of the elder Scipio Africanus and the first of the Caesars,
who were so-called because the uterus of his mother was cut [birth by C-
section]. It is for this reason also that the Caesones are so-called.*

V. COMPLETE PASSAGE FROM SECTION III

Herodotos *Historiae* 1.47.2-1.48.2

ὅ τι μὲν νυν τὰ λοιπὰ τῶν χρηστηρίων ἐθέσπισε, οὐ λέγεται πρὸς οὐδαμῶν: ἐν δὲ
Δελφοῖσι ὡς ἐσῆλθον τάχιστα ἐς τὸ μέγαρον οἱ Λυδοὶ χρησόμενοι τῷ θεῷ καὶ ἐπειρώτων
τὸ ἐντεταλμένον, ἡ Πυθίη ἐν ἑξαμέτρῳ τόνῳ λέγει τάδε. [3]

“οἶδα δ’ ἐγὼ ψάμμου τ’ ἀριθμὸν καὶ μέτρα θαλάσσης,
καὶ κωφοῦ συνήμι, καὶ οὐ φωνεῦντος ἀκούω.
ὀδμή μ’ ἐς φρένας ἦλθε κραταιρίνοιο χελώνης

ἐνομένης ἐν χαλκῷ ἅμ' ἀρνείοισι κρέεσσιν,
ἧ χαλκὸς μὲν ὑπέστρωται, χαλκὸν δ' ἐπιέσται.
ταῦτα οἱ Λυδοὶ θεσπισάσης τῆς Πυθίης συγγραψάμενοι οἶχοντο ἀπιόντες ἐς τὰς
Σάρδις. ὡς δὲ καὶ ἄλλοι οἰπεριπεμφθέντες παρήσαν φέροντες τοὺς χρησμούς, ἐνθαῦτα ὁ
Κροῖσος ἕκαστα ἀναπτύσσων ἐπώρα τῶν συγγραμμάτων, τῶν μὲν δὴ οὐδὲν προσίετό μιν·
ὁ δὲ ὡς τὸ ἐκ Δελφῶν ἤκουσε, αὐτίκα προσεύχετό τε καὶ προσεδέξατο, νομίσας μῶνον
εἶναι μαντήιον τὸ ἐν Δελφοῖσι, ὅτι οἱ ἐξευρήκεε τὰ αὐτὸς ἐποίησε. [2] ἐπεῖτε γάρδη
διέπεμψε παρὰ τὰ χρηστήρια τοὺς θεοπρόπους, φυλάξας τὴν κυρίην τῶν ἡμερέων
ἐμηχανᾶτο τοιάδε: ἐπινοήσας τὰ ἦν ἀμήχανον ἐξευρεῖν τε καὶ ἐπιφράσασθαι, χελώνην
καὶ ἄρνα κατακόψας ὁμοῦ ἤψε αὐτὸς ἐνλέβητι χαλκῷ, χάλκεον ἐπίθημα ἐπιθείς.

Now, on the one hand, what was prophesied by the rest of the oracles no one says. On the other hand, in Delphi, as soon as the Lydians went into the hall they consulted the god and asked what they were ordered to ask. The Pythian says the following in hexameter verse:

*I know the number of the grains of sand and the measure of the sea,
and I understand the mute and hear him who does not speak.
the smell has come to my mind of a strong-shelled tortoise
boiling in a cauldron together with the flesh of a lamb,
under which is bronze and bronze is above it.*

The Lydians, after the Pythian gave her prophecy, wrote down these things, departed, and went back into Sardis. When the others who were sent were present and brought their oracles, Croesus then unfolded each one and examined the writings. Some did not satisfy him at all. But when he heard [this one] from the Delphians, he immediately worshipped it and welcomed it, having decided that the oracle at Delphi was the only true one, since it had discovered the things that he himself did. For after he sent his envoys to the oracles, he devised the following and kept the date: he came up with something which was unable to be conjectured and unexpected – he cut up a tortoise and a lamb and he himself boiled them in a bronze cauldron and had covered it with a bronze lid.